Assignment I — Problem Set

Research in Finance

Prof. Dr. Andre Guettler Oliver Padmaperuma

Overview

> **Due:** 19 January 2026 (midnight) via email. **Submission format:** one `.R` script + one PDF (compiled from the Overleaf template on Moodle). **Group size:** up to **5** students. **Weight: 50%** of the final grade.

Solve the problem set posted on Moodle, building on the skills taught in Lectures 1–3 (R fundamentals, data handling, statistical analysis).

Submission rules

You must submit **only two files**:

1. **One `.R` script** — well-commented, self-explanatory, and efficient code (meaningful variable names, functions for repetitive tasks, no needless for-loops). Include text and comments inside the file that fully answer **both** Exercise 1 and Exercise 2.
2. **One PDF** that includes the outputs from your `.R` script (plots, calculations), the LaTeX tables built with `stargazer`, all plots with their captions and titles, and the written text following the **Overleaf template** posted on Moodle. **11 pt Times New Roman, 1.5 spaced.** The PDF is the polished, readable compilation of your results; the `.R` script provides the reproducible source.

Email the files to oliver.padmaperuma@uni-ulm.de, with andre.guettler@uni-ulm.de and your teammates in CC. Subject and file-naming pattern: `RiF2025_ProblemSet_surname1_surname2_...`. If attachment size is an issue, share a cloud link in the email body.

Exercise 1 — Prepare your data

Step	Task
a) Get the data	Create an `.R` script for all work and load libraries. Set your API key and load Quandl futures data as in class.[1] Merge positions and concentrations by date into `merged_gold`, `merged_silver`, `merged_btc`, `merged_eth`; drop NAs; sort each ascending by date.
b) Understand the data	Check the official sources (Nasdaq Data Link & CFTC[2]) so you can write 3–5 sentences for your *Data* section, explaining for novices what data you are using (weekly reports, trader types, etc.).
c) Join data sets	Combine `merged_gold`, `merged_silver`, `merged_btc`, and `merged_eth` into a data frame `combined` and create a variable `Asset` that captures from which original data set each observation came.
d) Clean & transform	Filter `combined` to dates after **1 April 2021** and save as `combined_clean`. Derive: *Net Long Position* per trader type;[3] *Weekly Change* for all *Net Long* positions (`week n / week n-1 - 1`); one variable each to designate *Year*, *Quarter*, and *Month*.
e) Descriptive analysis	Compute summary statistics[4] per asset and the *Net Long Position* / *Weekly Change* variables and create a table for the *Data* section.
f) Analysis & plot	Briefly discuss emerging patterns (e.g., higher crypto volatility) and propose plots that further visualise these patterns. Implement **at least two** corresponding `ggplot` plots and explain each in 4–5 sentences to enrich your *Data* section.

[1] Positions: `QDL/LFON`, `type = "FO_L_ALL"`. Concentrations: `QDL/FCR`, `type = "FO_L_ALL_CR"`. Contract codes: gold `088691`, silver `084691`, Bitcoin `133741`, Ethereum `146021`. Save as `com_gold` / `conc_gold`, etc.

[2] Nasdaq Data Link CFTC futures metrics: https://data.nasdaq.com/databases/CFTC; CFTC Commitments of Traders: https://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm.

[3] Commercial, non-commercial, total reportable, non-reportable, largest 4.

[4] Mean, standard deviation, minimum, 10th percentile, median, 90th percentile, maximum.

Exercise 2 — Write a paper

Use the **Overleaf template** on Moodle. Page minima are listed below for each section.

Step	Task	Min. length
a) Define a research question	Formulate a simple RQ that is testable with plots, calculations, or tests (e.g., *"Do crypto assets show higher Net Long Position volatility?"*) based on the patterns you identified in Exercise 1. Briefly motivate and explain in the *Introduction*.[5]	0.5 page
b) Conduct a literature review	Check whether there are 1–2 papers on which you can build, and briefly explain in 3–4 sentences.	0.5 page
c) Define your approach	Briefly describe how you are going to answer your RQ and why this approach makes sense.	0.5 page
d) Report insights	Report results in the *Results* section with at least **two** LaTeX tables built with `stargazer` (academic style) as well as at least **two** captioned plots.	1.5 pages
e) Explain & discuss	Summarise your results and discuss the implications of your research. *(Quality over length.)*	0.5 page

Sections in the Overleaf template: *Introduction*, *Literature Review*, *Methodology*, *Data* (1.5 pages), *Results* (1.5 pages), *Conclusion*.

---

[5] You may use an LLM (e.g., ChatGPT) **to refine** the RQ and the literature review — not to generate the substantive content. Disclose any AI assistance.

Grading rubric

Criterion	Weight
Code quality (efficiency, clarity, comments, naming)	30%
Creativity (plot design, RQ choice)	25%
Writing (concise, skim-friendly per the writing tips, with economic / practical justifications)	25%
Correctness of derivations and analyses	20%

Honor code

By submitting this assignment, you confirm that the work is your group's own, that all sources are cited, and that any AI assistance has been disclosed.